The Annals of Statistics
1997, Vol. 25, No. 5, 1954–1983

ABSTRACT TUBES, IMPROVED INCLUSION–EXCLUSION IDENTITIES AND INEQUALITIES AND IMPORTANCE SAMPLING

BY DANIEL Q. NAIMAN[1] AND HENRY P. WYNN[2]

Johns Hopkins University and City University, London

Numerous statistical applications require the evaluation of the probability content of a convex polyhedron. We demonstrate for a given polyhedron in R^d that there is a depth d inclusion–exclusion identity for its indicator function, which is a linear combination of indicator functions of intersections of at most d half-spaces. Terms in the identity are determined by the incidence of the facets of the polyhedron, which can be found using linear programming. This identity can be truncated at any depth to give a lower or upper bound. In addition, the resulting inequalities lead to importance sampling schemes for evaluating the probability content, and these methods tend to be more efficient than the naive hit-or-miss Monte Carlo method.

These results arise in a more general setting which we introduce. An *abstract tube* consists of a pair $(\mathscr{A}, \mathscr{S})$ where $\mathscr{A} = \{A_1, \dots, A_n\}$ is a collection of sets, $\mathscr{S}$ is a simplicial complex, and where each subcomplex $\mathscr{S}(x) = \{F \in \mathscr{S}: x \in \bigcap_{i \in F} A_i\}$ is contractible whenever $x \in \bigcup_{i=1}^n A_i$. The notion presented here is stronger than the one introduced earlier by Naiman and Wynn. Several examples are given and key consequences are demonstrated. In particular, arrangements of points and half-spaces in R^d give rise to abstract tubes via Voronoi decompositions and their associated Delauney dual complexes. Every abstract tube is shown to give rise to an inclusion–exclusion identity for $I_{\bigcup_{i=1}^n A_i}$, and upper and lower bounds are obtained by truncating the identity at an even or an odd depth. This property is analogous to the truncation inequality property of the classical inclusion–exclusion identity, which may be viewed as a special case. The notion of an abstract subtube is introduced, and it is shown that if $(\mathscr{A}, \mathscr{S}_1)$ is a subtube of $(\mathscr{A}, \mathscr{S}_2)$ then the truncation inequality gotten from the depth m truncation for $(\mathscr{A}, \mathscr{S}_1)$ is at least as sharp as the corresponding inequality from $(\mathscr{A}, \mathscr{S}_2)$. As a consequence, the generalized inclusion–exclusion inequalities are always at least as sharp as their classical counterparts.

1. Background. In Naiman and Wynn (1992), the authors introduced and discussed inclusion–exclusion identities for indicator functions of unions of Euclidean balls and spherical caps based on *Voronoi decompositions*. As in

Received April 1994; revised April 1997.
[1]Research partially supported by NSF Grant DMS-95-04242.
[2]Research partially supported by UK Engineering and Physical Sciences Research Council Grant GR/K81737.
AMS 1991 *subject classifications*. Primary 60E15, 62J01; secondary 62J05, 62F25, 60D05.
Key words and phrases. Tubes, inclusion–exclusion, simultaneous confidence intervals.

the case of the classical inclusion–exclusion identity [see Takács (1967)], these identities express the indicator function of a union as a linear combination of indicator functions of intersections. Except for very special cases, the identities are simpler in complexity than their classical inclusion–exclusion identity counterparts: under general position assumptions (described in Section 2.2.1), the depth of the identity, that is, the maximum number of intersections to consider, is $d + 1$, where d is the dimension of the ambient space. Naiman and Wynn (1992) described applications of these identities to multiple comparisons. Since then close connections have been exhibited with the theory of independence number and, through a point-set duality, to Vapnik–Chervonenkis dimension [see Naiman and Wynn (1993a, c)]. In this paper, we present a general framework in which the above examples arise as special cases, and we present some new developments which are critical for applying the resulting geometric methods to certain statistical problems.

In Section 2 we introduce the notion of an abstract tube and we present several examples. Briefly, an abstract tube consists of an indexed collection of sets and a simplicial complex (see Section 2 for the definition). Associated to any abstract tube is a generalized inclusion–exclusion identity for the indicator function of the union of the sets in the collection. For some of the examples of abstract tubes, for example, those in Naiman and Wynn (1992), where the sets are balls, the simplicial complex is based on a Voronoi decomposition. We present a new example of an abstract tube where the sets are half-spaces and the simplicial complex is obtained from a different Voronoi decomposition. In addition, we describe an example of an abstract tube associated with any finite collection of half-spaces, where the simplicial complex is obtained from the face incidence for the resulting polyhedron. This leads to inequalities for the probability content of the polyhedron and to a family of importance sampling schemes for this content.

In Section 3 we show that just as in the case of the classical inclusion–exclusion identity, there is a series of bounds that can be obtained from the basic Voronoi identity by truncating the identity at some depth. These alternate between being lower and upper bounds, depending on whether the truncation depth is odd or even. Furthermore, these inequalities are sharper than their classical counterparts. The proofs of these results rely on some simple techniques from algebraic topology. Notably, we use a purely algebraic topological result about chain complexes which forms the basis for the Morse inequalities of differential geometry.

Many statistical applications require inequalities for or evaluation of the Gaussian probability content of a convex polyhedron. Naiman and Wynn (1992), Section 5, describe how this problem arises in constructing simultaneous confidence intervals in ANOVA. Tong (1980) gives an extensive overview of probability bounds for the multivariate Gaussian probability of various types of sets. One consequence of the results in Section 3 is that we obtain new bounds for

$$P[(X_1, \dots, X_d) \in \mathscr{P}],$$

where $\mathscr{P} \subseteq R^d$ is a convex polyhedron. These bounds are of the form

$$(1) \qquad 1 - \sum_{k=0}^{m} (-1)^k \sum_{F \in \mathscr{F}^k} P\left[(X_1, \ldots, X_d) \in \bigcap_{i \in F} H_i^c\right],$$

where the H_i are the half-spaces defining the *facets* of $\mathscr{P}$, and where $\mathscr{F}^k$ denotes the collection of index sets F of size $k+1$ for which the bounding hyperplanes ∂H_i, $i \in F$ meet at a common point of the boundary of $\mathscr{P}$. For even m we get a lower bound and for odd m we get an upper bound. In particular, under a general position assumption if $(X_1, \ldots, X_d)$ has a Gaussian distribution, then the term

$$P\left[(X_1, \ldots, X_d) \in \bigcap_{i \in F} H_i^c\right]$$

involves the evaluations of an at most $\#(F)$-variate multivariate normal distribution function, where $\#(F)$ denotes the number of indices in F. For $k > 0$ the number of terms to evaluate at depth $k+1$, that is, the number of $k+1$-tuples in $\mathscr{F}^k$, is typically smaller than the corresponding number $\binom{n}{k+1}$ that is required by the classical inclusion–exclusion identity, where $n = \#(\mathscr{F}^0)$ is the number of facets of $\mathscr{P}$.

In Section 4 we describe how the inclusion–exclusion inequalities can be used to define importance sampling schemes. We describe in detail how these results can be used to evaluate the probability content of a polyhedron under a Gaussian distribution, and we compare these procedures numerically in some examples.

2. Abstract tubes.

2.1. *Terminology.* We proceed to introduce some relevant terminology and present a general framework in which all of our results apply. Many standard definitions from algebraic topology will be required, and some are given below. Standard algebraic topology texts [e.g., Rotman (1988)] can be consulted for further details.

Since this framework is somewhat abstract, the reader will probably benefit from a review of the classical inclusion–exclusion identity viewed as an indicator function identity, bearing in mind that our goal is generalization of it.

Given a finite collection of sets $\mathscr{A} = \{A_1, \ldots, A_n\}$, the classical inclusion–exclusion formula states

$$(2) \qquad I_{\bigcup_{i=1}^n A_i} = \sum_{\emptyset \neq J \subseteq \{1, \ldots, n\}} (-1)^{\#(J)-1} I_{\bigcap_{i \in J} A_i},$$

where I_A denotes the indicator function of the set A. In addition, when the right-hand side is truncated at some depth, a lower or upper bound is obtained, depending on whether truncation is at an odd or an even depth. A concise version of this statement is

$$(3) \qquad (-1)^m I_{\bigcup_{i=1}^n A_i} \le (-1)^m \sum_{J \subseteq \{1,\dots,n\},\ \#(J) \le m+1} (-1)^{\#(J)-1} I_{\bigcap_{i \in J} A_i},$$

for $m = 0, 1, 2, \dots$.

The right-hand side in (2) is a sum over collections of subsets of the index set $\{1, \dots, n\}$ and it is natural to ask whether there are smaller collections for which these same properties hold. For example, it is obvious that one need only sum over index sets J for which $\bigcap_{i \in J} A_i \neq \emptyset$. However, we are interested in more subtle improvements that arise from geometric structure of sets in the collection $\mathscr{A}$. Geometric considerations lead to our ability to make use of algebraic–topological tools, and in order to make use of these tools, we must restrict the collections of index sets to have the structure of a *simplicial complex*.

DEFINITION 1. A simplicial complex $\mathscr{S}$ is a collection of nonempty subsets of $\{1, \dots, n\}$ with the property that $F \in \mathscr{S}$ whenever $\emptyset \neq F \subseteq E$ and $E \in \mathscr{S}$. Elements $F \in \mathscr{S}$ with $\#(F) = k + 1$ are called k-dimensional faces or simplices. Zero-dimensional faces are called vertices, one-dimensional faces are called edges, and two-dimensional faces are called triangles. A subset of a simplicial complex that is itself a simplicial complex is called a subsimplicial complex.

A simplicial complex $\mathscr{S}$ is a combinatoric object, its simplices being finite sets. Such an object can be represented *geometrically* as a subset of Euclidean space by something called a *geometric realization*, constructed by identifying vertices of $\mathscr{S}$ with distinct points in a Euclidean space, connecting pairs of points corresponding to edges of $\mathscr{S}$, filling in triangles defined by triangles of $\mathscr{S}$, and so on. Thus, every face of $\mathscr{S}$ corresponds to an open geometric simplex in the geometric realization. Some care has to be taken in choosing the points that represent vertices in order to ensure that unwanted intersections do not occur. One device for doing this is to take the points corresponding to vertices of $\mathscr{S}$ to be the vertices of a sufficiently high-dimensional *geometric* simplex; however, these realizations are difficult to represent pictorially.

We will use the notation $|\mathscr{S}|$ to denote a geometric realization of $\mathscr{S}$. There is a slight problem with notation because there is a large set of geometric realizations associated with any particular $\mathscr{S}$. However, whatever we say about $|\mathscr{S}|$ will be true about any particular realization.

EXAMPLE 1. By the defining property, a simplicial complex can be specified by giving its maximal faces. Figure 1 shows a geometric realization of the simplicial complex whose maximal faces are $\{1, 2, 3\}$, $\{3, 4\}$, $\{4, 5\}$, $\{4, 6\}$ and $\{5, 6\}$.

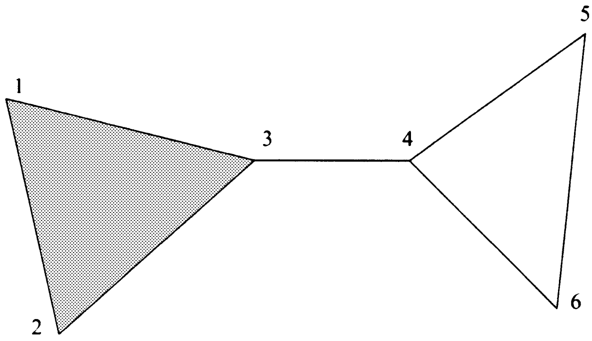

FIG. 1. *Geometric realization of a simplicial complex.*

EXAMPLE 2. The set of all nonempty subsets of $\{1, \dots, n\}$, that is, the index set summed over in (2) forms a simplicial complex that is geometrically realized by an $n - 1$-dimensional geometric simplex.

We now introduce the notion of an *abstract tube*, which is a modification of the original definition in Naiman and Wynn (1992). To proceed we need to discuss *contractibility* for a topological space. The formal definition says that a space is contractible if it is homotopy equivalent to a space consisting of a single point. Informally, homotopy equivalence between a pair of spaces means that the two spaces can be *deformed continuously* from one to the other. For example, a doughnut is homotopy equivalent to a coffee mug. In particular, contractibility of a space says that it can be squashed to a point by a continuous deformation, so that the space cannot contain any holes.

For example, a geometric simplex of any dimension is contractible. The *boundary* of a geometric simplex of positive dimension is not contractible. The geometric realization in Figure 1 is not contractible because of the hole enclosed by the edges of the $\{4, 5, 6\}$ triangle. Adding the interior of the $\{4, 5, 6\}$ triangle to the complex or removing one of its edges makes the space contractible. The geometric realization in Figure 2 *is* contractible.

We can refer to simplicial complexes as being homotopy equivalent if their geometric realizations are homotopy equivalent and we refer to a simplicial complex $\mathscr{S}$ as contractible if it has a contractible geometric realization.

DEFINITION 2. A pair $(\mathscr{A}, \mathscr{S})$ consisting of simplicial complex $\mathscr{S}$ whose vertices are among $\{1\}, \dots, \{n\}$ and an indexed collection $\mathscr{A} = \{A_i, i = 1, \dots, n\}$ of subsets of some set X is an abstract tube, provided that for every $x \in \bigcup_{i=1}^{n} A_i$ the subcomplex

$$\mathscr{S}(x) = \left\{F \in \mathscr{S} : x \in \bigcap_{i \in F} A_i\right\}$$

is contractible.

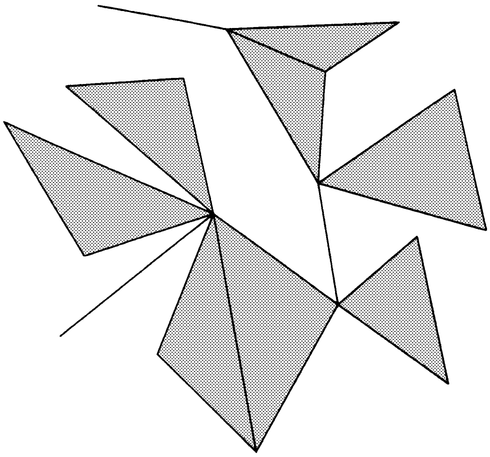

FIG. 2. *A contractible simplicial complex.*

EXAMPLE 3. Figure 3 gives an abstract tube formed by five disks $A_1, \dots, A_5$ and a simplicial complex $\mathscr{S}$. Each portion of $\bigcup_{i=1}^{5} A_i$ is labeled by the indices corresponding to the sets A_i that contain it. To check the abstract tube property, it is necessary to check contractibility of a subsimplicial complex corresponding to each piece. For example, for the point x_1, we have $\{i: x_1 \in A_i\} = \{2, 3, 4, 5\}$. The subsimplicial complex $\mathscr{S}(x_1)$, which consists of all of the faces of $\mathscr{S}$ formed from the index set $\{2, 3, 4, 5\}$, is shown in Figure 4, and is clearly contractible. Similarly, $\mathscr{S}(x_2)$ is a union of edges 12 and 24 in Figure 4 and is contractible. Note that every one of the regions labeled ij must correspond to an edge of $\mathscr{S}$, since otherwise the subsimplicial complex would be a two-point set, which is not contractible. The reader is invited to check the contractibility condition for the simplicial complex corresponding to each region in Figure 3.

EXAMPLE 4. Let $\mathscr{A} = \{A_i,\ i = 1, \dots, n\}$ be any finite collection of sets and let $\mathscr{S}$ denote the set of all subsets of $\{1, \dots, n\}$. As mentioned in Example 2, this simplicial complex is geometrically realized as an $n - 1$, dimensional simplex. For $x \in \bigcup_{i=1}^{n} A_i$ the subcomplex $\mathscr{S}(x)$ consists of all subsets of the nonempty set $J_x = \{j: x \in A_j\}$, which is geometrically realized as a $\#(J_x) -$ 1-dimensional simplex, so contractibility, and hence the abstract tube property, follows.

For convenience, we use $\mathscr{S}^k$ to denote the set of faces of $\mathscr{S}$ of dimension k. The *Euler characteristic* of a simplicial complex $\mathscr{S}$ is defined by

$$\chi(\mathscr{S}) = \sum_k (-1)^k \#(\mathscr{S}^k).$$

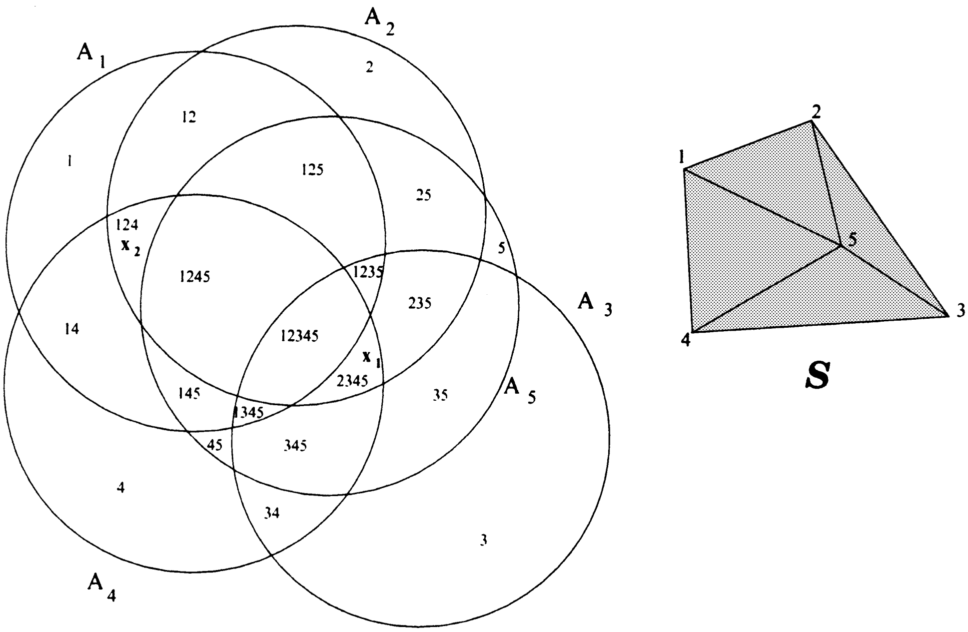

FIG. 3. *The five disks together with the simplicial complex shown form an abstract tube.*

This quantity is a topological invariant: if two simplicial complexes have homotopy equivalent geometric realizations, then they have the same Euler characteristic. In particular, if $\mathscr{S}$ is contractible, then $\chi(\mathscr{S}) = 1$.

It follows that for an abstract tube $(\{A_1, \dots, A_n\}, \mathscr{S})$ we have

$$I_{\bigcup_{i=1}^n A_i} = \sum_{F \in \mathscr{S}} (-1)^{\#(F)-1} I_{\bigcap_{i \in F} A_i}. \tag{4}$$

To see this, observe that the right-hand side, evaluated at some $x \in \bigcup_{i=1}^n A_i$ coincides with $\chi(\mathscr{S}(x))$, which is unity.

For the abstract tube of Example 4 from identity (4), we recover the classical inclusion–exclusion identity (2).

For completeness, we also mention that, in the context of Example 4, there is a dual abstract tube $(\mathscr{A}^*, \mathscr{S})$ where $\mathscr{A}^* = \{A_i^c,\ i = 1, \dots, n\}$ so that each A_i is replaced by its complement. For this case, (4) becomes

$$I_{\bigcup_{i=1}^n A_i^c} = \sum_{\varnothing \neq J \subseteq \{1, \dots, n\}} (-1)^{\#(J)-1} I_{\bigcap_{i \in J} A_i^c}; \tag{5}$$

alternatively,

$$I_{\bigcap_{i=1}^n A_i} = \sum_{\varnothing \neq J \subseteq \{1, \dots, n\}} (-1)^{\#(J)-1} I_{\bigcup_{i \in J} A_i}. \tag{6}$$

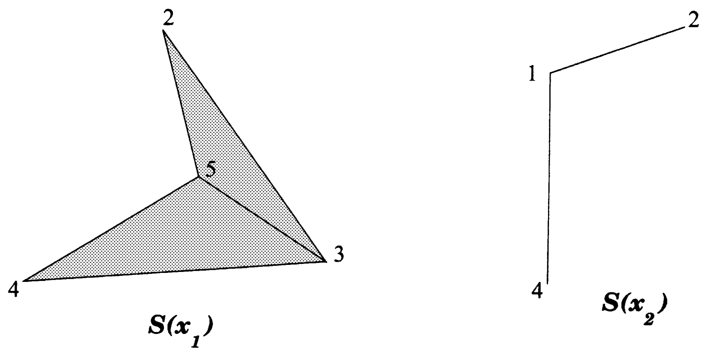

FIG. 4. *The subsimplicial complexes of Example* 2.

Identity (4) appears in Naiman and Wynn (1992) as Theorem 2.1. There the contractibility requirement does not appear in the definition of an abstract tube; only the weaker condition $\chi(\mathscr{S}(x)) = 1$ was used. The stronger contractibility assumption will be used in Section 3 to prove inequalities, and identity (4) arises as a special case (see Theorem 4).

REMARK 1. The definition of an abstract tube can be weakened by requiring contractibility of $\mathscr{S}(x)$ for almost every x with respect to some dominating measure μ on X. In this case, the pair $(\mathscr{A}, \mathscr{S})$ might be referred to as a *weak abstract tube* with respect to μ. For every weak abstract tube, we obtain an identity (4) which holds almost everywhere with respect to μ.

DEFINITION 3. An abstract tube $(\mathscr{A}_1, \mathscr{S}_1)$ is a *subtube* of the abstract tube $(\mathscr{A}_2, \mathscr{S}_2)$ provided that $\mathscr{S}_1$ is a subsimplicial complex of $\mathscr{S}_2$ and $\mathscr{A}_1 = \mathscr{A}_2$.

REMARK 2. In some cases of interest (see Sections 2.2.1 and 2.2.2) there is an indexed family of abstract tubes $(\mathscr{A}_t, \mathscr{S})$ where the same simplicial complex appears while the collection $\mathscr{A}_t = \{A_i(t),\ i = 1, \dots, n\}$ of sets varies with the index t. This observation has an important statistical motivation. A typical goal is the determination of critical points for some procedure, and this may translate into determining a critical value for the parameter t so that the probability content of $\bigcup_{i=1}^{n} A_i(t)$ is bounded above or below by some prescribed value. The procedure for finding this critical value is much simplified if the simplicial complex giving identity (4) or a related inequality remains fixed for all values of t.

2.2. *Abstract tubes via nerves*. For the more interesting abstract tubes described below, the simplicial complex arises as a *nerve*, and the proof of the abstract tube property makes use of a remarkable duality theorem of Borsuk (1948). We review the relevant concepts here. In the following, we consider coverings of a subset of Euclidean space by polyhedra, where a polyhedron is defined as an intersection of closed half-spaces. Borsuk's theorem concerns

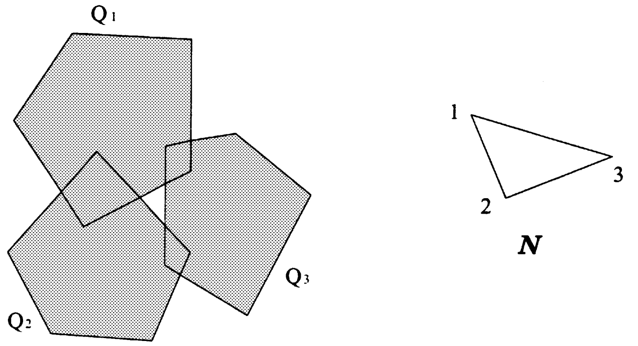

FIG. 5. *Borsuk's theorem: the nerve of a covering is homotopy equivalent to the space covered.*

coverings of a space having the regularity property that each intersection of sets in the covering is either empty or what is called an *absolute retract*, which is more general than what we need, so we state a simpler version of this theorem that applies in all of the cases considered below.

Let $C = \{Q_1, \dots, Q_n\}$ be a covering of a set X; that is, $X = \bigcup_{i=1}^{n} Q_i$. The nerve defined by C is the simplicial complex

$$\mathscr{N}(C) = \left\{ J \subseteq \{1, \dots, n\} \colon J \neq \varnothing \text{ and } \bigcap_{i \in J} Q_i \neq \varnothing \right\}.$$

THEOREM 1 [Borsuk (1948)]. *If $C = \{Q_1, \dots, Q_n\}$ is a covering of X by polyhedra in R^d then X and $|\mathscr{N}(C)|$ are homotopy equivalent.*

EXAMPLE 5. Figure 5 gives an example to illustrate nerves and Borsuk's theorem. Here the union of Q_1, Q_2 and Q_3 forms a space that is homotopy equivalent to the simplicial complex $\mathscr{N}$ shown, which is the nerve defined by the Q_i.

2.2.1. *Abstract tube associated with a polyhedron.* This example, which is inspired by an observation of Edelsbrunner (1993), Section 5, leads to an inclusion–exclusion identity for the indicator function of a polyhedron which will be used in Section 4.3.

Let $\mathscr{P}$ be a d-dimensional convex polyhedron in R^d with exactly n facets (the dimension of a facet is $d - 1$) denoted $F_1, \dots, F_n$. (See Figure 6). Let H_i be the closed half-space with $F_i \subseteq H_i$ and for which the bounding hyperplane is a support hyperplane to $\mathscr{P}$ corresponding to F_i so that $F_i = \mathscr{P} \cap \partial H_i$ and $\mathscr{P} = \bigcap_{i=1}^{n} H_i$.

Let $\mathscr{F}$ be the simplicial complex defined by the nerve of the covering $\{F_1, \dots, F_n\}$ of the boundary of $\mathscr{P}$. That is,

$$\mathscr{F} = \left\{ J \subseteq \{1, \dots, n\} \colon J \neq \varnothing \text{ and } \bigcap_{i \in J} F_i \neq \varnothing \right\},$$

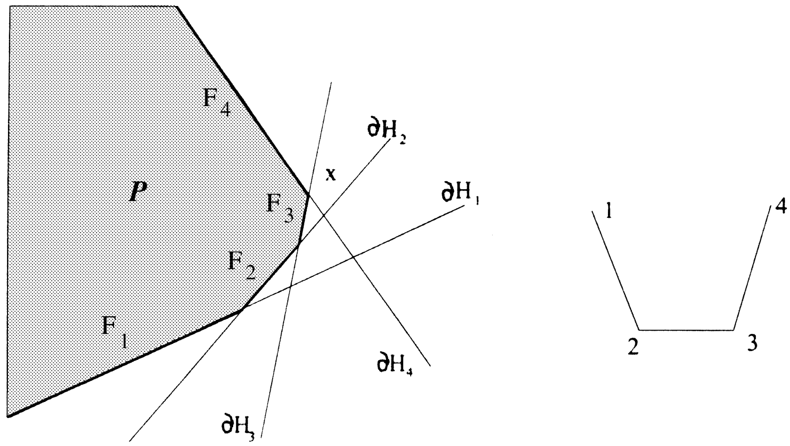

FIG. 6. *The simplicial complex associated with a polyhedron.*

so that this simplicial complex encodes the *facet incidences* for $\mathscr{P}$. In addition, we take

$$\mathscr{H} = \{ H_i^c : i = 1, \dots, n \}.$$

THEOREM 2. *The pair* $(\mathscr{H}, \mathscr{F})$ *forms an abstract tube.*

PROOF. Fix $x \in \bigcup_{i=1}^{n} H_i^c = \mathscr{P}^c$. We must show $\mathscr{F}(x)$, is contractible. In fact, Edelsbrunner (1993) notes that $\mathscr{F}(x)$ consists of all of the faces (including lower-dimensional ones) that are *visible* from x, and he proves by an induction that this subcomplex has Euler characteristic. It is not difficult to strengthen this conclusion and give an alternative proof of his result.

Let $J_x = \{i \colon x \in H_i^c\}$. Then $\{F_i,\ i \in J_x\}$ is the set of facets of $\mathscr{P}$ visible from x. The union of these facets $U = \bigcup_{i \in J_x} F_i$ forms the portion of the boundary of $\mathscr{P}$ that is visible from x. For example, for the point x shown in Figure 6, the facets visible from x are F_3 and F_4, so $U = F_3 \cup F_4$, which is a union of two line segments having one common end point. In general, the set U is contractible by the following argument.

Let K be a hyperplane separating x from $\mathscr{P}$. Let $\pi \colon \partial\mathscr{P} \to K$ map each $y \in \mathscr{P}$ to the intersection of the ray $\overline{xy}$ with K. The restriction of π to U defines a homeomorphism between U and $\pi(U)$, and $\pi(U)$ is easily seen to be convex, hence contractible. On the other hand, $\mathscr{F}(x)$ is the nerve of $\{F_i,\ i \in J_x\}$ which is a covering of U by polyhedra, so by Theorem 1, $\mathscr{F}(x)$ and U are homotopy equivalent, and we conclude that $\mathscr{F}(x)$ is contractible. □

We say a d-dimensional polyhedron $\mathscr{P}$ is in *general position* if the corresponding simplicial complex $\mathscr{F}$ is at most d-dimensional; that is, if there are no collections of $d + 1$ incident facets. If $\mathscr{P}$ fails to be in general position, it can be perturbed slightly (in possibly several different ways), as shown in Lemma 1, to give a polyhedron $\tilde{\mathscr{P}}$ that *is* in general position. In addition, it will be shown in Corollary 1 that the pair consisting of the collection of complementary half-spaces for the original polyhedron $\mathscr{P}$, and the simplicial complex associated with the perturbed polyhedron $\tilde{\mathscr{P}}$, forms a weak abstract tube.

Let $\mathscr{P} \subseteq R^d$ be a d-dimensional polyhedron. Assume the origin is in the interior of $\mathscr{P}$ (which may require a translation) so that we have $\mathscr{P} = \bigcap_{i=1}^n H_i$, where

$$H_i = \left\{x \in R^d \colon u_i^t x \leq 1\right\},$$

for some $u_i \in R^d \setminus \{0\}$, and each H_i intersects $\mathscr{P}$ in a distinct facet. Let $\mathscr{F}$ denote the simplicial complex associated with $\mathscr{P}$.

For $\varepsilon \in R^n$ and $\delta \in R$, let $\mathscr{P}^{\varepsilon,\delta} = \bigcap_{i=1}^n H_i^{\varepsilon,\delta}$ denote the *perturbed* polyhedron, with bounding half-spaces

$$H_i^{\varepsilon,\delta} = \left\{x \in R^d \colon u_i^t x \leq 1 + \delta\varepsilon_i\right\}.$$

LEMMA 1. (a) *For any $\varepsilon \in R^d \setminus \{0\}$, there exists $\Delta = \Delta(\varepsilon) > 0$ such that for $\delta \in (0, \Delta)$ we have the following.*

(i) *Each of the sets $\mathscr{P}^{\varepsilon,\delta} \cap H_i^{\varepsilon,\delta}$, $i = 1, \dots, n$, is a facet of the polyhedron $\mathscr{P}^{\varepsilon,\delta}$, so that the associated simplicial complex $\mathscr{F}^{\varepsilon,\delta}$ has the same vertex set $\{1, \dots, n\}$ as $\mathscr{F}$;*

(ii) *The simplicial complex $\mathscr{F}^{\varepsilon,\delta}$ is a subcomplex of $\mathscr{F}$;*

(iii) *$\mathscr{F}^{\varepsilon,\delta^*} = \mathscr{F}^{\varepsilon,\delta}$ for all $\delta^* \in (0, \delta)$.*

(b) *There exists a set $E \subseteq R^n$, which is a union of at most finitely many linear subspaces of dimension less than n, with the property that for any $\varepsilon \in R^n \setminus E$ the polyhedron $\mathscr{P}^{\varepsilon,\delta}$ is in general position, for $\delta \in (0, \Delta(\varepsilon))$.*

The proof of Lemma 1 is somewhat long and appears in the Appendix.

COROLLARY 1. *If $\mathscr{H} = \{H_1^c, \dots, H_n^c\}$, then the pair $(\mathscr{H}, \mathscr{F}^{\varepsilon,\delta})$ forms a weak abstract tube with respect to Lebesgue measure, that is, a subtube of $(\mathscr{H}, \mathscr{F})$, with $\mathscr{F}^{\varepsilon,\delta}$ at most d-dimensional, provided $\varepsilon \in R^n \setminus E$, and $\delta \in (0, \Delta(\varepsilon))$, where E and $\Delta(\varepsilon)$ are as given in Lemma 1.*

PROOF. Let $\mathscr{H}^{\varepsilon,\delta} = \{(H_1^{\varepsilon,\delta})^c, \dots, (H_n^{\varepsilon,\delta})^c\}$, so that the pair $(\mathscr{H}^{\varepsilon,\delta}, \mathscr{F}^{\varepsilon,\delta})$ is the abstract tube associated with the polyhedron $\mathscr{P}^{\varepsilon,\delta}$. The fact that $(\mathscr{H}, \mathscr{F}^{\varepsilon,\delta})$ is a subtube of $(\mathscr{H}, \mathscr{F})$ follows from Lemma 1(a)(ii), and the dimensionality bound follows from (b), so we only need to verify the weak abstract tube property for $(\mathscr{H}, \mathscr{F}^{\varepsilon,\delta})$. To this end, define $B = \bigcup_{i=1}^n \partial H_i^c$, which is a set of Lebesgue measure 0 in R^d. Fix $x \in \bigcup_{i=1}^n H_i^c \setminus B$. We proceed to demonstrate contractibility of the subcomplex $\{J \in \mathscr{F}^{\varepsilon,\delta} \colon x \in \bigcap_{i \in J} H_i^c\}$. For δ sufficiently small we have $\{i \colon x \in H_i^c\} = \{i \colon x \in (H_i^{\varepsilon,\delta})^c\}$, so that $x \in \bigcup_{i=1}^n (H_i^{\varepsilon,\delta})^c$ and

$$\left\{J \in \mathscr{F}^{\varepsilon,\delta} \colon x \in \bigcap_{i \in J} \left(H_i^{\varepsilon,\delta}\right)^c\right\} = \left\{J \in \mathscr{F}^{\varepsilon,\delta} \colon x \in \bigcap_{i \in J} H_i^c\right\}.$$

Using the fact that $(\mathscr{H}^{\varepsilon,\delta}, \mathscr{P}^{\varepsilon,\delta})$ is an abstract tube, the desired contractibility property follows. □

If $\mathscr{P}$ is in general position then the corresponding inclusion–exclusion identity (4) has the attractive feature that there are no terms in the sum on the right-hand side involving more than d-fold intersections, and the total number of terms is at most $\binom{n}{d}$, which is a polynomial in n. This is in contrast with the classical inclusion–exclusion identity (2), which has exponentially many $(2^n - 1)$ terms.

If $\mathscr{P}$ is not in general position, any one of the weak abstract tubes $(\mathscr{H}, \mathscr{F}^{\varepsilon,\delta})$ from Corollary 1 gives rise to an inclusion–exclusion identity (4) for $I_{\mathscr{P}}$ that holds a.e. with respect to Lebesgue measure, and whose complexity is the same as what is obtained for a polyhedron in general position. In addition, Theorem 5 and the subtube property guarantee that the abstract tube $(\mathscr{H}, \mathscr{F}^{\varepsilon,\delta})$ gives rise to truncation inequalities that are at least as sharp as those obtained from $(\mathscr{H}, \mathscr{F})$.

Whether or not $\mathscr{P}$ is in general position, an algorithm to determine whether a given index $J \subseteq \{1, \dots, n\}$ is in the associated simplicial complex $\mathscr{F}$ is easy to construct using linear programming. Again, assume $H_i = \{x \in R^d: u_i^t x \le 1\}$, where each u_i is a nonzero d-vector. Then $\bigcap_{i \in J} F_i \ne \emptyset$ is equivalent to the feasibility of the following linear system:

$$
\text{(7)} \qquad \begin{aligned} u_i^t x &= 1 \quad \text{for } i \in J, \\ u_i^t x &\le 1 \quad \text{for } i \notin J. \end{aligned}
$$

A standard technique for deciding this feasibility question [see Murty (1983)] is to solve the following linear program:

$$
\text{(8)} \qquad \begin{array}{lll} \text{minimize} & z & \\ \text{subject to} & u_i^t x - u_i^t y + z = 1 & \text{for } i \in J, \\ & u_i^t x - u_i^t y + z \le 1 & \text{for } i \notin J, \\ & x \ge 0 & \\ & y \ge 0 & \\ & z \ge 0. & \end{array}
$$

Here, inequality is taken to be componentwise. This system is feasible since we can take $z = 1$, $x = 0$ and $y = 0$. The system (7) is feasible if and only if the optimal solution (x, y, z) to the system (8) satisfies $z = 0$.

The results in Sections 3 and 4 give inequalities that can be derived from *skeletons* of $\mathscr{F}$ of various dimensions. Here, the q-skeleton $\mathscr{F}_q$ of $\mathscr{F}$ is defined as the subsimplicial complex of $\mathscr{F}$ consisting of those faces whose dimension is at most q. Calculation of $\mathscr{F}_q$ is computationally simpler than calculation of $\mathscr{F}$ for small values of q, since one need only check feasibility of each system (7) where J ranges throughout the set of all $\binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{q+1}$ subsets of $\{1, \dots, n\}$ of cardinality at most $q + 1$. In particular, determining the pairs of incidence faces in $\mathscr{F}$ involves checking feasibility for $O(n^2)$ systems (7).

2.2.2. *Arrangements of balls.* This example is a review and slight generalization of the result in Naiman and Wynn (1992). See also Edelsbrunner (1993).

DEFINITION 4. An arrangement of balls is an indexed collection of sets $\mathscr{B}_t = \{B_i(t),\ i = 1,\dots,n\}$ for some fixed $t \in R$, where

$$B_i(t) = \left\{x \in R^d : \|x - x_i\|^2 \le s_i + t\right\}$$

for distinct points $x_i \in R^d$ and constants s_i.

The term *arrangement of balls* actually refers here to a *family* of collections of sets indexed by t. It might seem more natural to define an arrangement of balls as above, taking $t = 0$. However, it turns out that when we use Definition 4, we can introduce a single simplicial complex that makes every $\mathscr{B}_t$, $t \in R$ into an abstract tube, so Remark 2 becomes relevant. This acknowledges the possibility that some of the balls in an arrangement for a particular value of t might be empty. On the other hand, each ball will be nonempty for t sufficiently large.

An arrangement $\mathscr{B}_t$ leads to a decomposition of R^d into a union of Voronoi polyhedra meeting only on their boundaries. These polyhedra are defined by

$$V_j = \left\{x \in R^d : \|x - x_j\|^2 - s_j = \min_{1 \le i \le n} \|x - x_i\|^2 - s_i\right\}. \tag{9}$$

Edelsbrunner (1993) refers to $d_j(x) := \|x - x_j\|^2 - s_j$ as the *power distance* from x to the ball $B_j(0)$.

To define a simplicial complex, we use the *nerve* construction to yield the so-called Delauney dual simplicial complex. Thus, the vertices of $\mathscr{D}$ are taken to be the $\{j\}$ for which $V_j \neq \varnothing$, and the faces of $\mathscr{D}$ are the nonempty index sets J for which $\bigcap_{j \in J} V_j \neq \varnothing$. Observe that the vertex set of $\mathscr{D}$ may be a proper subset of $\{1,\dots,n\}$. Properties of the Voronoi decomposition and Delauney dual are discussed in Naiman and Wynn (1992) and Edelsbrunner (1986, 1987, 1993).

To prove that $(\mathscr{B}_t, \mathscr{D})$ forms an abstract tube, we follow the same argument as in the proof of Lemma 3.1 and Theorem 3.1 of Naiman and Wynn (1992). Fix a point $x \in \bigcup_{i=1}^n B_i(t)$ and define $J_x = \{i\colon x \in B_i(t)\}$. The analogue of Lemma 3.1 says that for $x \in \bigcup_{i=1}^n B_i(t)$, the set $\bigcup_{i \in J_x} V_i$ is star-shaped with respect to x, hence contractible. In the modified proof, we simply replace the definition of $S(x_i, x_k)$ by

$$S(x_i, x_k) := \left\{x \in R^d : d_i(x) \le d_k(x)\right\}.$$

For the abstract tube property, we follow the proof technique sketched on page 59, last paragraph, of Naiman and Wynn (1992). We have

$$\mathscr{D}(x) = \{I \in \mathscr{D} : I \subseteq J_x\},$$

which coincides with the *nerve* of the covering $\{V_i,\ i \in J_x\}$ of $\bigcup_{i \in J_x} V_i$. Since the sets V_i are polyhedra, Theorem 1 applies and $\bigcup_{i \in J_x} V_i$ and $\mathscr{D}(x)$ have the same homotopy type, so contractibility of $\mathscr{D}(x)$ follows from that of $\bigcup_{i \in J_x} V_i$.

The identity (3.3) in Naiman and Wynn (1992) is (4) specialized to this case. Note that $(\mathscr{B}_t, \mathscr{D})$, in general, forms an abstract subtube of the abstract tube defined using the simplicial complex consisting of all subsets of the vertex set, as in Example 4. When these tubes coincide, the Voronoi method does not lead to an improvement to the inclusion–exclusion identity. In fact, the two tubes coincide if and only if there exists a point $x \in R^d$ whose power distances from all of the $B_i(0)$ is the same. If all the s_i are equal, this condition says that there exists a point x equidisant from the x_i (in the Euclidean sense) so that the x_i lie on a common sphere.

2.2.3. *Dual Voronoi diagrams.* Using the same notation as in Section 2.2.2, define the collection of complementary balls $\mathscr{B}_t^* = \{B_i(t)^c,\ i = 1, \dots, n\}$. A *furthest distance* Voronoi decomposition is obtained when min is replaced by max in (9). The nerve construction then leads to a simplicial complex which will be denoted by $\mathscr{D}^*$.

The same argument as is used for $(\mathscr{B}_t, \mathscr{D})$ shows that $(\mathscr{B}_t^*, \mathscr{D}^*)$ forms an abstract tube.

The resulting identity (4) is an identity for the indicator function of $\bigcup_{i=1}^n B_i(t)^c$, equivalently, an identity for $\bigcap_{i=1}^n B_i(t)$. The relationship between the identities gotten from $(\mathrm{B}_t, \mathscr{D})$ and $(\mathscr{B}_t^*, \mathscr{D}^*)$ is completely analogous to the relationship between $(\mathscr{A}, \mathscr{S})$ of Example 4 and $(\mathscr{A}^*, \mathscr{S})$ used to give (5). Some remarkable properties of these identities are given in Naiman and Wynn (1993b).

2.2.4. *Half-spaces.*

Definition 5. A half-space arrangement is an indexed collection of half-spaces

$$\mathscr{H}_t = \{H_i(t),\ i = 1, \dots, n\},$$

where

$$H_i(t) = \{x \in R^d : \langle x, u_i \rangle \le c_i + t\},$$

for some distinct unit d-vectors u_i and constants c_i and t.

For such an arrangement, we define a Voronoi decomposition by taking

$$\tilde{V}_i = \left\{x \in R^d : \langle x, u_i \rangle - c_i = \min_{1 \le j \le n} \langle x, u_j \rangle - c_j\right\}$$

so that

$$\tilde{V}_i = \bigcap_{j=1}^{n} \tilde{S}(i|j),$$

where

$$\tilde{S}(i|j) = \{x \in R^d : \langle x, u_i \rangle - c_i \le \langle x, u_j \rangle - c_j\}.$$

We define the analogue of the Delauney simplicial complex

$$\tilde{\mathscr{D}} = \left\{J \subseteq \{1, \dots, n\} : \bigcap_{i \in J} \tilde{V}_i \ne \varnothing\right\}.$$

Again, there is a dual Voronoi decomposition for half-spaces defined by replacing min by max so that

$$\tilde{V}_i^* = \bigcap_{j=1}^{n} \tilde{S}(j|i),$$

and we denote the corresponding Delauney simplicial complex by $\tilde{\mathscr{D}}^*$. Determining if a given index set is a simplex of $\tilde{\mathscr{D}}$ (or its dual) reduces to checking feasibility of a certain linear program. As in the case of balls we define

$$\mathscr{H}_t^* = \left\{H_i(t)^c,\, i = 1, \dots, n\right\}.$$

It is easy to verify that the Voronoi decomposition (dual Voronoi decomposition) for the half-space arrangement defined by (u_i, c_i) coincides with the dual Voronoi decomposition (Voronoi decomposition) for the ball arrangement with $(x_i, r_i) = (u_i, 1 - 2c_i)$. Furthermore, the Voronoi decomposition for the half-space arrangement defined by (u_i, c_i) coincides with the dual Voronoi decomposition for the half-space arrangement defined by $(-u_i, -c_i)$, and vice versa.

A completely analogous argument to the one used for arrangements of balls yields the conclusion that the pairs $(\mathscr{H}_t, \tilde{\mathscr{D}})$ and $(\mathscr{H}_t^*, \tilde{\mathscr{D}}^*)$ form abstract tubes. In particular, there is a critical star-shaped property analogous in statement and proof to Lemma 3.1 in Naiman and Wynn (1992). This says that for $t \in R$ and $x \in \bigcup_{i=1}^n H_i(t)$, if the set $\bigcup_{i \in J_x} \tilde{V}_i$ is nonempty, where $J_x = \{i\colon x \in H_i(t)\}$ then it is star-shaped with respect to x.

2.2.5. *Spherical caps.* Let S^{d-1} denote the unit $d-1$-sphere in R^d, and define the angular distance between $u, v \in S^{d-1}$ by $d(u,v) = \cos^{-1}\langle u, v\rangle$. For fixed points u_i, $i = 1, \dots, n \in S^{d-1}$, the $\theta \in (0, \pi/2)$ define a collection of spherical caps

$$\mathscr{C} = \{C_i,\, i = 1, \dots, n\},$$

where

$$C_i = \left\{v \in S^{d-1}\colon d(v, u_i) \le \theta\right\}.$$

Define a *spherical* Voronoi decomposition by taking

$$V_i = \left\{v \in S^{d-1}\colon d(v, u_i) \le d(v, u_j) \text{ for all } j = 1, \dots, n\right\}$$

and let $\mathscr{D}$ be the corresponding nerve.

The result in Naiman and Wynn (1992), Theorem 4.1, can be interpreted as saying that $(\mathscr{C}, \mathscr{D})$ forms an abstract tube if $\bigcap_{i=1}^n C_i = \varnothing$.

3. Abstract tube identities and truncation inequalities. In this section we consider truncated versions of the expressions on the right-hand side of (4) and prove that they give upper or lower bounds for the left-hand side. From a practical standpoint, these bounds can be enormously useful because, typically, lower-depth terms are easier to evaluate. For example, in evaluating the Gaussian probability content of a polyhedron, the sets A_i are half-spaces, and depth 1 terms reduce to univariate normal integrals, depth 2 terms to bivariate normal integrals, and so on.

The proof of the basic truncation inequality makes use of the following elementary fact which, incidentally, forms the basis for the Morse inequalities of differential geometry [see Hirsch (1976)]. It is interesting to note that while the Morse inequalities relate information about the number of critical points of various indices for a Morse function on a manifold, they are based on the following more fundamental algebraic topological result. While this latter result is implicit in treatments of the Morse inequalities, it is surprising to us that this result is not found in standard algebraic topology textbooks.

The approach we take here employs some elementary homology theory. For the definition of simplicial homology, we recommend the exposition in Hilton (1960). For a simplicial complex $\mathscr{S}$, we define a chain complex

$$\cdots C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} C_{n-2} \xrightarrow{\partial_{n-2}} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} C_{-1} \cdots, \tag{10}$$

where C_i is a free Abelian group whose rank is γ_i, the number of simplices of dimension i of $\mathscr{S}$, and the homology groups of this chain complex coincide with the usual singular homology groups of geometric realizations $|\mathscr{S}|$. By convention, we take $C_n = (0)$ for $n < 0$. Let b_i denote the rank of the image B_i of C_{i+1} under ∂_{i+1} and let z_i denote the rank of the kernel Z_i of ∂_i, for the chain complex (10). Here the term *rank* refers to the number of copies of $\mathbb{Z}$ that appear when the group is decomposed, as a direct sum of cyclic groups. We define $\dim(\mathscr{S}) = \max\{i: \gamma_i \neq 0\}$.

THEOREM 3. *Let $\mathscr{S}$ be a simplicial complex with γ_i simplices of dimension i, for $i = 0, 1, \dots$. Let β_i denote the ith Betti number of $\mathscr{S}$; that is, β_i denotes the rank of the ith simplicial homology group, that is, the quotient group $H_i = Z_i/B_i$. Then*

$$\sum_{i=0}^{m} (-1)^i \beta_i = \sum_{i=0}^{m} (-1)^i \gamma_i + (-1)^{m+1} b_m \tag{11}$$

for $m = 0, 1, \dots$.

The first term on the right hand side of (11) may be recognized as the Euler characteristic of the m-skeleton of $\mathscr{S}$. For the case when $m = \dim(\mathscr{S})$, we have $b_m = 0$ and we obtain the well-known conclusion that the alternating sum of the Betti numbers gives the Euler characteristic of $\mathscr{S}$.

PROOF. The proof is reminiscent of the proof presented in introductory algebraic topology courses of the fact that the Euler characteristic is a topological invariant. From the short exact sequences,

$$0 \to Z_n \xrightarrow{i} C_n \xrightarrow{\partial_n} B_{n-1} \to 0, \qquad n = 0, 1, 2, \dots$$

we obtain

$$z_n + b_{n-1} = \gamma_n, \qquad n = 0, 1, 2, \dots.$$

On the other hand, the exact sequence that defines H_n as a quotient group

$$0 \to B_n \xrightarrow{i} Z_n \to H_n \to 0, \qquad n = 0, 1, 2, \dots,$$

leads to

$$b_n + \beta_n = z_n,\ n = 0, 1, \dots .$$

Equation (11) follows easily. □

Since $b_i \geq 0$ always, and since a contractible simplicial complex has Betti numbers $\beta_i = 1$ if $i = 0$ and $\beta_i = 0$, for $i > 0$ we obtain the following.

COROLLARY 2. *If $\mathscr{S}$ is a finite simplicial complex then we have the inequalities*

$$(-1)^m \sum_{i=0}^{m} (-1)^i \beta_i \leq (-1)^m \sum_{i=0}^{m} (-1)^i \gamma_i$$

for $m = 0, 1, \dots,$ with equality if $m = \dim(\mathscr{S})$. In particular, if $\mathscr{S}$ is contractible then

$$(-1)^m \leq (-1)^m \sum_{i=0}^{m} (-1)^i \gamma_i,$$

for $m = 0, 1, 2, \dots$ and equality holds if $m = \dim(\mathscr{S})$.

Returning to abstract tubes, if $\mathscr{S}$ is a simplicial complex, recall that we use $\mathscr{S}_m$ to denote the m-skeleton of $\mathscr{S}$, and $\mathscr{S}^k$ to denote the set of k-dimensional simplices of $\mathscr{S}$ so that

$$\mathscr{S}_m = \bigcup_{i=0}^{m} \mathscr{S}^i.$$

THEOREM 4. *For any abstract tube $(\mathscr{A}, \mathscr{S})$ with $\mathscr{A} = \{A_i,\ i = 1, \dots, n\}$, we have*

$$(-1)^m I_{\bigcup_{i=1}^n A_i} \leq (-1)^m \sum_{J \in \mathscr{S}_m} (-1)^{\#(J)-1} I_{\bigcap_{i \in J} A_i}, \tag{12}$$

for $m = 0, 1, 2, \dots,$ with equality if $m = \dim(\mathscr{S}) = \max\{\#(J)\colon J \in \mathscr{S}\} - 1$.

Note that for $m = 0$ we obtain the familiar Bonferroni bound, while for $m = \dim(\mathscr{S})$ we obtain identity (4).

PROOF. If $x \notin \bigcup_{i=1}^n A_i$, both sides of (12) vanish. On the other hand, if $x \in \bigcup_{i=1}^n A_i$ then the left-hand side of (12) evaluated at x gives $(-1)^m$ and the right-hand side equals $(-1)^m \sum_{i=0}^m (-1)^i \gamma_i$, where $\gamma_i = \#(\mathscr{S}(x)^i)$, the number of simplices of $\mathscr{S}(x)$ of dimension i. Thus, the result follows from Corollary 2. □

As a special case, using the abstract tubes defined in Example 4, we recover the classical inclusion–exclusion identity (2) and the associated truncation inequalities (3). Furthermore, from the abstract tube associated with a convex polyhedron (see Section 2.2.1) we obtain the inequalities in (1).

THEOREM 5. *If* $(\mathscr{A},\mathscr{S})$ *forms an abstract subtube of* $(\mathscr{A},\mathscr{S}^*)$ *then*

$$(13)\quad (-1)^m \sum_{J\in\mathscr{S}m} (-1)^{\#(J)-1} I_{\bigcap_{i\in J} A_i} \le (-1)^m \sum_{J\in\mathscr{S}^*_m} (-1)^{\#(J)-1} I_{\bigcap_{i\in J} A_i}$$

for $m = 0, 1, 2, \dots$.

REMARK 3. Since any abstract tube is a subtube of the abstract tube defined as in Example 4, we may conclude from this result that every inequality obtained by truncation of an abstract tube identity (4) at depth m is at least as sharp as the inequality obtained by truncation of the classical inclusion–exclusion identity (2) at depth m. The problem of finding sharpest abstract tube inequalities reduces to the interesting combinatorial problem of finding the minimal abstract tubes.

PROOF OF THEOREM 5. Both sides of (13) evaluated at $x \notin \bigcup_{i=1}^n A_i$ are zero, so assume $x \in \bigcup_{i=1}^n A_i$. The chain complex (10) for $\mathscr{S}(x)$ gives Abelian groups C_i, B_i, Z_i and H_i whose corresponding ranks are denoted by γ_i, b_i, z_i and β_i. Similarly, let β_i^*, b_i^*, z_i^* and γ_i^* be the ranks of the groups C_i^*, B_i^*, Z_i^* and H_i^* corresponding to $\mathscr{S}^*(x)$. Evaluating both sides of (13) at x we see that it suffices to show

$$(-1)^m \sum_{i=0}^m (-1)^i \gamma_i \le (-1)^m \sum_{i=0}^m (-1)^i \gamma_i^*.$$

By the contractibility of $\mathscr{S}(x)$ [and $\mathscr{S}^*(x)$], we have $\beta_i^* = \beta_i$, $i = 0, 1, \dots$. Using Theorem 3 we see that

$$(-1)^m \sum_{i=0}^m (-1)^i \gamma_i = (-1)^m \sum_{i=0}^m (-1)^i \gamma_i^* + b_m - b_m^*.$$

Since the chain complex $C_.$ can be viewed as a subcomplex of the chain complex $C^*_.$, B_i can be viewed as a subgroup of B_i^* for every i, and so $b_m \le b_m^*$ for $m = 0, 1, \dots,$ and we obtain (13). □

4. Importance sampling.

4.1. *Use of inclusion–exclusion inequalities to devise importance sampling schemes*. The truncation inequalities in Section 3 lead to efficient importance sampling schemes for determining the probability content of a union of sets. Fix an abstract tube $(\{A_1, \dots, A_n\}, \mathscr{S})$ with $A_i \in \mathscr{G}$, a σ-algebra of subsets of X, a probability measure P on $(X, \mathscr{G})$, and suppose our goal is to approximate $P(\bigcup_{i=1}^n A_i)$.

Let $m \geq 0$ and assume that we are able to evaluate the upper bound

$$(-1)^{m-1} \sum_{J \in \mathscr{S}_{m-1}} (-1)^{\#(J)-1} P\left[\bigcap_{i \in J} A_i\right], \tag{14}$$

for

$$(-1)^{m-1} P\left(\bigcup_{i=1}^{n} A_i\right) \tag{15}$$

obtained by truncation of the identity (4) at depth m (i.e., the inequality based on the $m-1$-skeleton of $\mathscr{S}$). When $m = 0$ we define $\mathscr{S}_{-1} = \varnothing$ so that the bound in (14) (for $-P(\bigcup_{i=1}^{n} A_i)$) is 0. Recall that $\mathscr{S}^i$ denotes the set of simplices of $\mathscr{S}$ of dimension i, while $\mathscr{S}_m$ denotes the set of simplices of dimension at most m.

Define a probability density function

$$f_m = \frac{1}{Q_m} \sum_{J \in \mathscr{S}^m} I_{\bigcap_{i \in J} A_i},$$

on X with respect to P; where

$$Q_m = \sum_{J \in \mathscr{S}^m} P\left[\bigcap_{i \in J} A_i\right],$$

let

$$h_m = I_{\bigcup_{i=1}^{n} A_i} - \sum_{J \in \mathscr{S}_{m-1}} (-1)^{\#(J)-1} I_{\bigcap_{i \in J} A_i},$$

and let $g_m = h_m / f_m$ in the support of f_m. The main result of this section is the following.

THEOREM 6. *Let $Y_1, \dots, Y_k$ be iid and distributed according to the probability distribution of $f_m\, dP$; then*

$$\sum_{J \in \mathscr{S}_{m-1}} (-1)^{\#(J)-1} P\left(\bigcap_{i \in J} A_i\right) + \sum_{i=1}^{k} g_m(Y_i)/k$$

is an unbiased estimator for $P[\bigcup_{i=1}^{n} A_i]$.

PROOF. The idea is to use importance sampling to approximate the difference between (14) and (15). Using the depth $m+1$ truncation inequality we see that

$$\begin{aligned} (-1)^m I_{\bigcup_{i=1}^{n} A_i} &\leq (-1)^m \sum_{J \in \mathscr{S}_m} (-1)^{\#(J)-1} I_{\bigcap_{i \in J} A_i} \\ &= (-1)^m \sum_{J \in \mathscr{S}_{m-1}} (-1)^{\#(J)-1} I_{\bigcap_{i \in J} A_i} + \sum_{J \in \mathscr{S}^m} I_{\bigcap_{i \in J} A_i}, \end{aligned}$$

so combining with the depth m inequality, we obtain

$$0 \leq (-1)^m \left[I_{\bigcup_{i=1}^n A_i} - \sum_{J \in \mathscr{S}_{m-1}} (-1)^{\#(J)-1} I_{\bigcap_{i \in J} A_i} \right] \leq \sum_{J \in \mathscr{S}^m} I_{\bigcap_{i \in J} A_i}.$$

Equivalently,

$$0 \leq (-1)^m h_m \leq Q_m f_m.$$

It follows that

$$\int_X h_m \, dP = \int_{f_m > 0} h_m \, dP = \int_{f_m > 0} \frac{h_m}{f_m} f_m \, dP = E_{fm}[g_m(Y)]$$

and the result follows. □

REMARK 4. In Theorem 6 we obtain the same conclusion if f_m is replaced by $f'_m = \psi f_m$ and g_m is replaced by $g'_m = g_m/\psi$, where ψ is a nonnegative function, normalized so that ψf_m defines a probability density function with respect to P, and $\psi > 0$ on the set where $g_m \neq 0$.

4.2. *Discussion.* It is instructive to describe this importance sampling procedure for the easiest cases. Since it is typically the case that lower depth terms are less numerous and more tractable, we view the size of m as indicative of the level of complexity of the procedure.

The simplest case for this sampling procedure is when $m = 0$, so that we are determining the *correction* to the probability estimate of 0. The depth 1 inclusion–exclusion truncation bound utilizes the vertices, that is, the singletons $\{i\} \in \mathscr{S}$. These singletons are used to define a density function

$$f_0 = \frac{\sum_{\{i\} \in \mathscr{S}} I_{A_i}}{\sum_{\{i\} \in \mathscr{S}} P(A_i)}$$

from which we sample to obtain $Y_1, \ldots, Y_k$ and estimate $P[\bigcup_{i=1}^n A_i]$ using

$$\sum_{i=1}^k g_0(Y_i)/k,$$

where

$$g_0 = \frac{\sum_{\{i\} \in \mathscr{S}} P[A_i]}{\sum_{\{i\} \in \mathscr{S}} I_{A_i}}.$$

Note that the denominator of g_0 counts the number of A_i, $\{i\}$ a vertex of $\mathscr{S}$, containing a given point.

This special case is the analogue of the importance sampling procedure of Frigessi and Vercellis (1984) [see also Fishman (1996), Section 4.1.2] for the union counting problem. Here one is given finite sets $S_1, \ldots, S_n \subseteq S$, of known size and one wants to estimate $\#(\bigcup_{i=1}^n S_i)$. Let $Y_1, \ldots, Y_k$ be iid with each Y_i obtained as follows. Take $I_1, \ldots, I_n$ to be iid random indices with $P[I_i = j] \propto \#(S_j)$. Conditional on I_i, take Y_i to be uniformly distributed in

S_I, then estimate $\#(\bigcup_{i=1}^{n} S_i)$ using $(\#(S)/k)\sum_{j=1}^{k}(1/\#\{i\colon Y \in \mathscr{S}_i\}$. This estimator is known to outperform the naive hit or miss estimator, where Y_i are randomly sampled from S and the estimator is the proportion of Y_i in $\bigcup_{i=1}^{n} S_i$, in cases when the union is small relative to S. A detailed analysis appears in Fishman (1996).

For $m = 1$, we are correcting the Bonferroni bound $\sum_{\{i\} \in \mathscr{S}} P[A_i]$, which uses the *vertices* of $\mathscr{S}$, by sampling from the density function that is defined using the *edges* of $\mathscr{S}$. If $Y_1, \dots, Y_k$ is a sample from the distribution with density

$$f_1 = \frac{\sum_{\{i,j\} \in \mathscr{S}} I_{A_i \cap A_j}}{\sum_{\{i,j\} \in \mathscr{S}} P\left[A_i \cap A_j\right]}$$

(with respect to P) then we estimate $P[\bigcup_{i=1}^{n} A_i]$ using

$$\sum_{\{i\} \in \mathscr{S}} P[A_i] + \sum_{j=1}^{k} g_1(Y_j)/k,$$

where

$$g_1 = \sum_{\{i,j\} \in \mathscr{S}} P\left[A_i \cap A_j\right] \frac{I_{\bigcup_{=1}^{n} A_i} - \sum_{\{i\} \in \mathscr{S}} I_{A_i}}{\sum_{\{i,j\} \in \mathscr{S}} I_{A_i \cap A_j}}.$$

4.3. *Importance sampling for the Gaussian probability content of a polyhedron.* We now summarize how Theorem 6 combined with the abstract tube of Section 2.2.1 can be used to approximate the Gaussian probability content of a polyhedron. As in Section 2.2.1, assume a d-dimensional polyhedron $\mathscr{P}$ is given and we want to approximate $P[Z \in \mathscr{P}]$ where $Z \sim N_d(\mu, \Sigma)$. Assume $\mathscr{P} = \bigcap_{i=1}^{n} H_i \subseteq R^d$, where the facets of $\mathscr{P}$ are given by $F_i = H_i \cap \mathscr{P}$. Assume $H_i = \{x \in R^d\colon u_i^t x \le 1\}$, where each u_i is a nonzero d-vector.

We describe the approximation method as the following sequence of steps. Throughout the discussion we assume vectors are represented as columns. Fix a depth $m \ge 0$. The *depth m* procedure is as follows.

Step 1. Determine the collection of index sets

$$\mathscr{S}_m = \left\{J \subseteq \{1, \dots, n\} \colon \#(J) \le m + 1, \bigcap_{i \in J} F_i \# \varnothing\right\}.$$

This step can be carried out using linear programming as described in the remarks at the end of Section 2.2.1.

Step 2. Calculate

$$p = \sum_{J \in \mathscr{S}_{m-1}} (-1)^{\#(J)-1} P\left(Z \in \bigcap_{i \in J} H_i^c\right).$$

Each probability in this sum is of the form

$$P\left[u_i^t Z \ge 1, i \in J\right], \tag{16}$$

where $\#(J) \le m$, so involves at most m-fold integration.

Step 3. Calculate each probability

$$q_J = P\left[Z \in \bigcap_{i \in J} H_i^c\right], \qquad J \in \mathscr{S}^m$$

which is of the form in (16) with $\#(J) = m + 1$, and so involves at most $m + 1$-fold integration.

Let $q = \sum_{J \in \mathscr{S}^m} q_J$ and use the q_J to express the distribution $f_m\, dF_Z$ as a mixture of distributions

$$f_m\, dF_Z = \sum_{J \in \mathscr{S}^m} \frac{q_J}{q} f_m^J\, dF_Z,$$

where

$$f_m^J = \frac{1}{q_J} I_{I \cap_{i \in J} H_i^c}.$$

Step 4. For each $J = \{i_1, \dots, i_{m+1}\} \in \mathscr{S}^m$, order the indices in J so that $\{u_{i_1}, \dots, u_{i_{k_J}}\}$ is a maximal linearly independent set of vectors. Then we have

$$u_{i_{k_J + p}} = \sum_{j=1}^{k_J} \lambda_{p,j}^J u_{i_j}, \qquad p = 1, \dots, m + 1 - k_J,$$

for some choice of constants $\lambda_{p,j}^J$. Let λ_p^J denote the column k_J-vector whose entries are the $\lambda_{p,j}$, $j = 1, \dots, k_J$.

Next find a basis $v_1, \dots, v_d$ of R^d with $v_j = u_{i_j}$, for $j = 1, \dots, k_J$. Let A_J denote the matrix whose rows are the v_i^t and define $W = AZ$. Write $W = \left[\begin{smallmatrix} W^{(1)} \\ W^{(2)} \end{smallmatrix}\right]$ where $W^{(1)}$ is k_J-dimensional and $W^{(2)}$ is $(d - k_J)$-dimensional. Find (column) vectors $\mu_J^{(1)}$, $\mu_J^{(2)}$, and matrices B_J, $\Sigma_J^{(1)}$ and $\Sigma_J^{(2)}$ such that (marginally) $W^{(1)} \sim N_{k_J}(\mu^{(1)}, \Sigma_J^{(1)})$ and conditionally, given $W^{(1)} = w^{(1)}$, $W^{(2)}$ is distributed as $N_{d-k_J}(\mu_J^{(2)} + B_J w(1), \Sigma^{(2)})$. Also, compute A_J^{-1}.

Using these quantities, we can sample from the probability distribution $f_m^J\, dF_Z$ as follows:

1. Generate $W^{(1)}$ having a $N_{k_J}(\mu^{(1)}, \Sigma_J^{(1)})$ distribution, conditional on $W_i^{(1)} \geq 1$, for $i = 1, \dots, k_J$, and $\lambda_p^{Jt} W^{(1)} \geq 1$, for $p = 1, \dots, m + 1 - k_J$.
2. Generate $W^{(2)}$ distributed according to $N_{d-k_J}(\mu_J^{(2)} + B_J W^{(1)}, \Sigma_J^{(2)})$.
3. Take $Z = A_J^{-1} W$.

Step 5. Draw a sample $Y_1, \dots, Y_k$ from $f_m\, dF_Z$ and estimate $P[Z \in \mathscr{P}]$ using

$$1 - p - \sum_{i=1}^{k} g_m(Y_i)/k,$$

where

$$g_m = \frac{I_{\cup_{i=1}^n H_i^c} - \sum_{J \in \mathscr{S}_{m-1}} (-1)^{\#(J)-1} I_{\cap_{i \in J} H_i^c}}{f_m}.$$

The conditional sampling in (1) warrants additional discussion. Typically, we are interested in the case when μ and Σ are both relatively small, so that the boundaries of the polytope are in the tails of the distribution of Z. For $m = 0$, in (1) we are sampling in the tails of a univariate normal distribution. Procedures for this purpose are well known and appear in Devroye (1986), for example. For $m \geq 1$ one idea for (1) is to generate $W^{(1)}$ by enclosing the set

$$O_J = \left\{x \in R^{m+1}: x_i \geq 1,\, i = 1,\dots,k_J,\, \lambda_p^{Jt} x \geq 1,\, p = 1,\dots,m+1-k_J\right\}$$

in a half-space H, then generate $W^{(1)}$ according to the conditional distribution of $W^{(1)}$ given $W^{(1)} \in H$, using the same technique as described in Step 4, until we get $W \in O_J$. Thus, sampling to get $W^{(1)} \in H$ essentially involves generating a random variate T in the tails of a univariate normal distribution, generating an m-variate normal random variate conditionally on T, then linear transforming. We have found this procedure to perform adequately in numerical examples when $m = 1$ and we leave the problem of carrying out this step more efficiently as an open problem.

REMARK 5. The method can be applied when Z is distributed as a discrete mixture of normal distributions so that conditionally, given $I = i$, we have $Z \sim N_d(\mu_i, \Sigma_i)$. All that is required is the incorporation of an additional step of repeated generation of I and evaluating each conditional probability $P[Z \in \mathscr{P} \mid I = i]$ as described.

REMARK 6. The method can also be used to approximate $P[Z \in \mathscr{P}]$ when the density function of Z is of the form $\Sigma_k\, r_k(z)\phi_k(z)$, where r_k is a known function and ϕ_k is a normal density. For example, Edgeworth expansions are of this form with each r_k being a polynomial. Here we approximate each term

$$\int_{\mathscr{P}} r_k(z)\,\phi_k(z)\,dz,$$

by utilizing Remark 4 as follows. Steps 1 and 4 remain unchanged. In Step 2, take

$$p = \sum_{J \in \mathscr{S}_{m-1}} (-1)^{\#(J)-1} \int_{\bigcap_{i \in J} H_i^c} r_k(z)\,\phi_k(z)\,dz,$$

and in Step 3 take

$$q_J = \int_{\bigcap_{i \in J} H_i^c} \phi_k(z)\,dz.$$

Finally, modify the function g_m in Step 5 by multiplying by the factor r_k.

REMARK 7. Steps 1 and 4 can require a significant amount of computation effort. On the other hand, it is frequently the case that we are interested in determining $P[Z \in c\mathscr{P}]$, where c takes a range of values. For example, in many testing situations we look for a critical value c so that this probability takes some given value $1 - \alpha$. In such cases, Step 1 needs to be carried out only once, while the quantities in Step 4 are easily modified to account for the change in c.

4.4. *Numerical examples*. In this section we present some examples to illustrate the procedure and describe some general phenomena that appear to deserve further study.

We have written C code for implementing the procedure described above for general problems. The linear programming was carried out using *lp_solve*, a public-domain linear programming package (with source code) written by Michel Berkelaar and Jeroen Dirks. The linear algebra and normal distribution function calculations utilized the *Numerical Recipes* in *C* source code provided by Press, Flannery, Toukolsky and Vetterling (1988). Finally, the bivariate normal distribution function was calculated using an algorithm due to Donnelly (1973).

The calculations described below were carried out on a SUN Sparcstation 5.

4.4.1. *A signal detection problem*. Consider the problem of detecting a *localized signal* in the sequence X_t, $t = 1, \dots, 20$, meaning that some local average of the X_t is sufficiently large. As a simple example, we take detection to mean that at least one of the following events occurs:

$$X_t > c \quad \text{for some } t = 1, \dots, 20,$$

$$X_t + X_{t+1} > \sqrt{2}\, c \quad \text{for some } t = 1, \dots, 19,$$

$$X_t + X_{t+1} + X_{t+2} > \sqrt{3}\, c \quad \text{for some } t = 1, \dots, 18,$$

where $c > 0$. We are interested in the probability of false detection, that is, detection of a signal under a suitable null hypothesis. For illustrative purposes we compute the probability p, of detection, under the white noise hypothesis, when X_t are iid $N(0, 1)$. Note that failure of detection corresponds to the vector $X = [X_1, \dots, X_{20}]$ lying in a convex polytope.

We compare three Monte Carlo sampling methods for calculating p, the probability of signal detection: the naive hit-or-miss method and the two importance sampling methods described above for depths 0 and 1. In order to compare the three methods, it is important to note that while all of them produce unbiased estimates based on averages of iid random variates, there is a *nesting* of the computational effort required in using the importance sampling procedures. Some overhead, namely, the determination of incidences (Step 1) and the matrix calculations (Step 4) can be carried out once and for all; this applies to the calculation of the whole table of probabilities as we vary the critical constant c. Additional overhead is required for each particular value of c, since there are multivariate tail probabilities to be determined in Steps 2 and 3. Further sampling effort could be carried out to reduce the variance of the estimates and the overhead costs would remain fixed.

TABLE 1
Sampling effort (CPU seconds) in Step 4 to obtain $\sigma_{\hat{p}} = p/10$ for the detection probability estimate $\hat{p}$

Method	Detection probability p					
	0.25	**0.10**	**0.05**	**0.01**	**0.005**	**0.001**
hit-or-miss	0.418	1.15	2.39	12.3	26.1	109
depth 0	0.072	0.0511	0.0418	0.0283	0.0244	0.0178
depth 1	0.404	0.147	0.0768	0.0242	0.0155	0.00640

Table 1 summarizes the sampling effort required by the three methods, ignoring the overhead just described. Since one might typically be interested in obtaining a prescribed relative error for a given estimator $\hat{p}$ of p, we estimated the amount of CPU time required to attain $\sigma_{\hat{p}} = p/10$, for each estimator. These calculations were performed as follows. For each of a range of critical values c, we used a sample of $N = 10{,}000$ variates and estimated the variance $\sigma^2_{\hat{p}_N}$ for each of the three estimators $\hat{p}_N$. In addition, we recorded the CPU time T_N required to carry out each sampling effort. By treating the average of all three estimates as the true value of p, we are able to solve the equation

$$\sigma_{\hat{p}_M} = \sqrt{N\sigma^2_{\hat{p}_N}/M} = p/10,$$

for M, to estimate the required sample size. Since the CPU time is proportional to the sample size, we can then estimate the required CPU time as $T_M = M(T_N/N) = 100\sigma^2_{\hat{p}_N}/p^2$.

We see from Table 1 that, as expected, the sampling effort required for the hit-or-miss procedure grows as p decreases, while the importance sampling procedures achieve the desired accuracy in progressively less time. In addition, in terms of sampling effort, the depth 1 procedure out performs the depth 0 procedure for sufficiently small values of p.

Table 2 gives the overhead required by the two importance sampling procedures. Since the computational effort required by Steps 2 and 3 varied only slightly for different values of p, the ranges are given. Note that the

TABLE 2
Overhead (CPU seconds) for depth 0 and depth 1 importance sampling procedures to obtain for the detection probability problem for p in the range in Table 1

	All values of p		Fixed value of p
Method	**Step 1 (incidences)**	**Step 4 (matrices)**	**Steps 2 and 3 (probabilities)**
depth 0	1.96	22.40	0.022 ± 0.001
depth 1	43.54	658.27	1.52 ± 0.01

overhead required to carry out the importance sampling procedures can be quite substantial relative to the sampling effort. Also, the depth 1 procedure seems to require an order of magnitude more computational effort than the depth 0 procedure in overhead.

We find that for computing a single probability (e.g., a p-value) that is sufficiently small ($p < 0.005$) the effort required by the depth 0 procedure, including overhead, is comparable to or less than the effort required to use the hit-or-miss procedure. By extrapolation, it appears that for $p < 0.0001$ the depth 1 procedure, which concentrates its effort in overhead, outperforms the hit-or-miss procedure. On the other hand, unless the overhead for the depth 1 procedure can be reduced, the depth 0 procedure appears more practical. Future investigations will focus on finding ways to control the overhead of the importance sampling procedures in special situations.

4.4.2. *Tukey–Kramer procedure*. This problem is discussed in Naiman and Wynn (1992). We do some calculations for a specific example. We consider a one-way ANOVA with eight cells. Let n_i denote the number of observations in the ith cell. We take $n_1, \dots, n_8$ to be 5, 6, 8, 10, 12, 13, 14, 16. By conditioning on the estimated error sum of squares, the problem of finding the coverage probability for simultaneous confidence intervals for all pairwise differences of cell means reduces to finding the probability that

$$|\hat{\mu}_i - \hat{\mu}_j| \le C\sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \quad \text{for all } i \ne j,$$

where $\hat{\mu}_i$ are independent $N(0,1)$ random variables.

Tables 3 and 4 give the same information as in Tables 1 and 2 for the Tukey–Kramer coverage probability calculations. We see the same qualitative features in this case as were pointed out for the signal detection problem. We believe that the phenomena observed, namely, improvement using higher depth inequalities sufficiently far out in the tails of the distribution, is quite general, and future investigation will attempt to better understand this observation.

TABLE 3

Sampling effort (CPU seconds) in Step 4 to obtain $\sigma_{\hat{p}} = p/10$ for the Tukey–Kramer coverage probability $\hat{p}$

	Detection probability p					
Method	**0.25**	**0.10**	**0.05**	**0.01**	**0.005**	**0.001**
hit-or-miss	0.144	0.425	0.851	4.03	8.06	29.7
depth 0	0.0256	0.0157	0.0110	0.00569	0.00424	0.00210
depth 1	0.100	0.0227	0.00787	0.00112	0.000516	0.0000789

TABLE 4

Overhead (CPU seconds) for depth 0 and depth 1 importance sampling schemes to obtain the Tukey–Kramer coverage probability p in the range in Table 3

	All values of p		**Fixed value of p**
Method	**Step 1 (incidences)**	**Step 4 (matrices)**	**Steps 2 and 3 (probabilities)**
depth 0	1.72	0.913	0.007 ± 0.001
depth 1	35.1	26.6	0.35 ± 0.03

APPENDIX

Proof of Lemma 1. To prove Lemma 1 we first need the following.

LEMMA A.1. *For any $\varepsilon \in R^d \setminus \{0\}$, if $\mathscr{P} \cap \partial H_i$ is a facet of $\mathscr{P}$ then there exists $\Lambda_i = \Lambda_i(\varepsilon) > 0$ such that $\mathscr{P}^{\varepsilon,\delta} \cap \partial H_i^{\varepsilon,\delta}$ is a facet of $\mathscr{P}^{\varepsilon,\delta}$ for all $\delta \in (0, \Lambda_i)$.*

PROOF. Fix a point x_i in the relative interior of $\mathscr{P} \cap \partial H_i$, that is, a solution to the system

$$u_i^t x_i = 1$$

$$u_j^t x_i < 1 \quad \text{for all } j \neq i.$$

Let $x_i^\delta = x_i + (\delta\varepsilon_i/\|u_i\|^2)u_i$ so that $u_i^t x_i^\delta = 1 + \delta\varepsilon_i$. If we define

$$\Lambda_i = \min\left\{\frac{1 - u_j^t x_i}{\varepsilon_i u_j^t u_i/\|u_i\|^2 - \varepsilon_j} : j \neq i \text{ and } \varepsilon_i u_j^t u_i/\|u_i\|^2 - \varepsilon_j > 0\right\},$$

then $\Lambda_i > 0$ and for $\delta \in (0, \Lambda_i)$ we have

$$u_j^t x_i^\delta = u_j^t x_i + \frac{\delta\varepsilon_i}{\|u_i\|^2} u_j^t u_i < 1 + \delta\varepsilon_j \quad \text{for all } j \neq i;$$

that is, x_i^δ lies in the relative interior of the facet $\mathscr{P}^{\varepsilon,\delta} \cap \partial H_i^{\varepsilon,\delta}$ of $\mathscr{P}^{\varepsilon,\delta}$. $\square$

LEMMA A.2. *For any $\varepsilon \in R^n \setminus \{0\}$, if $J \notin \mathscr{F}$ there exists $\Sigma_J = \Sigma_J(\varepsilon) > 0$ such that $J \notin \mathscr{F}^{\varepsilon,\delta}$ for all $\delta \in [0, \Sigma_J)$.*

PROOF. Fix $\varepsilon \in R^n \setminus \{0\}$, and $J \notin \mathscr{F}$, and let $m = \#(J)$. The system

$$\begin{aligned} u_i^t x &= 1, \quad i \in J, \\ u_i^t x &\leq 1, \quad i \notin J, \end{aligned} \tag{17}$$

is infeasible and we need to show that the system

$$\begin{aligned} u_i^t x &= 1 + \delta\varepsilon_i, \quad i \in J, \\ u_i^t x &\leq 1 + \delta\varepsilon_i, \quad i \notin J, \end{aligned} \tag{18}$$

is infeasible for $\delta > 0$ sufficiently small. By standard techniques from linear programming [see Murty (1983)] [similar to the reduction from (7) to (8)], the system (17) is equivalent to a system of the form

$$\begin{aligned} Ay &= b, \\ y &\geq 0, \end{aligned} \tag{19}$$

where A is an $(n+m) \times (2d+1)$ matrix depending only on the u_i and J, and b is an $n+m$ vector depending only on J. The same construction applied to (18) gives a system equivalent to (18) of the form

$$\begin{aligned} Ay &= b + \delta\varepsilon^*, \\ y &\geq 0, \end{aligned} \tag{20}$$

where ε^* is a nonzero vector depending only on ε and J.

Let $W^\delta = \{y \in R^{2d+1}\colon\ Ay = b + \delta\varepsilon^*\}$, and let $V = \{Ay\colon\ y \in R^{2d+1}\} \subseteq R^{n+m}$, the column space of A. Note that by assumption (19) is infeasible. Also, infeasibility of (20) holds if $b + \delta\varepsilon^* \notin V$ or equivalently, if $W^\delta = \varnothing$.

If $W^0 = \varnothing$ then $b \notin V$, so there exists an open neighborhood of b contained in V^c. It follows that $b + \delta\varepsilon^* \notin V$ for δ sufficiently small. If, on the other hand, $W^0 \neq \varnothing$, then there are two cases to consider. Either $\varepsilon^* \notin V$ in which case $b + \delta\varepsilon^* \notin V$, for all $\delta \neq 0$, by linearity of V, or $\varepsilon^* \in V$. For the latter case, if we let $A\tilde{y} = \varepsilon^*$, then $W^\delta = W^0 + \delta\tilde{y}$, for all $\delta \in R$. It follows that

$$d(W^\delta, W^0) = \inf\{\|x - y\|\colon x \in W^\delta,\ y \in W^0\} \leq \delta\|\tilde{y}\|.$$

Let $Q = \{y \in R^{2d+1}\colon\ y \geq 0\}$ denote the nonnegative orthant. Since, by assumption, (19) is infeasible, we have $W^0 \cap Q = \varnothing$, and $d(W^0, Q) > 0$, since these are closed polyhedra. By the triangle inequality we see that

$$d(W^\delta, Q) \geq d(W^0, Q) - d(W^\delta, W^0) > d(W^0, Q) - \delta\|\tilde{y}\| > 0,$$

and hence (20) is infeasible, if δ is sufficiently small. $\square$

LEMMA A.3. *For any* $\varepsilon \in R^n \setminus \{0\}$, *if* $J \in \mathscr{F}$, *there exists* $\Gamma_J = \Gamma_J(\varepsilon) > 0$ *such that either* $J \notin \mathscr{F}^{\varepsilon,\delta}$ *for all* $\delta \in (0, \Gamma_J)$, *or* $J \in \mathscr{F}^{\varepsilon,\delta}$ *for all* $\delta \in (0, \Gamma_J)$.

PROOF. As in the proof of Lemma A.2, we convert the condition for $J \in \mathscr{F}$ or $J \in \mathscr{F}^{\varepsilon,\delta}$ into a statement about feasibility of the systems (19) and (20). By assumption, (19) is feasible and we want to show that the feasibility or infeasibility of (20) is unchanged for $\delta > 0$ in a neighborhood of 0. Using the same notation as in the proof of Lemma A.2, we have $b \in V$ and $W^0 \neq \varnothing$.

If $\varepsilon^* \notin V$, then $b + \delta\varepsilon^* \notin V$ for all $\delta \neq 0$, by linearity. It follows that (20) is infeasible for all $\delta \neq 0$.

If $\varepsilon^* \in V$, then letting $A\tilde{y} = \varepsilon^*$ as above, we have $W^\delta = W^0 + \delta\tilde{y}$. Since (19) is feasible we have $W^0 \cap Q \neq \varnothing$. If $W^0 \cap Q^{\text{int}} \neq \varnothing$ then clearly $W^\delta \cap Q^{\text{int}} \neq \varnothing$; hence, (20) is feasible, for δ sufficiently small. If $W^0 \cap Q^{\text{int}} = \varnothing$

then by linearity of W^0, and the fact that Q is an orthant, we must have $0 \in W^0$. If (20) is feasible for some $\delta^* > 0$, let $y \in W^{\delta^*} \cap Q$. Then we have $y = \hat{y} + \delta^* \tilde{y}$ for some $\hat{y} \in W^0$. If $\lambda \in [0,1]$ convexity of W^0 and Q gives

$$\lambda y = \lambda \hat{y} + (\lambda\delta^*)\tilde{y} \in \left(W^0 + (\lambda\delta^*)\tilde{y}\right) \cap Q = W^{\lambda\delta^*} \cap Q,$$

we conclude that (20) is feasible for all $\delta \geq 0$ sufficiently small. □

Proof of Lemma 1. To prove (a) fix $\varepsilon \in R^d \setminus \{0\}$ and define $\Lambda = \min_{1 \le i \le n} \Lambda_i$, $\Sigma = \min_{J \notin \mathscr{F}} \Sigma_J$, and $\Gamma = \min_{J \in \mathscr{F}} \Gamma_J$, where Λ_i, Σ_J and Γ_J are defined in Lemmas A.1, A.2 and A.3. If we define $\Delta = \min\{\Lambda, \Sigma, \Gamma\}$ then clearly $\Delta > 0$ and properties (i), (ii), and (iii) in the statement of the lemma follow.

To prove (b), let $J_1, \dots, J_q$ be the faces of $\mathscr{F}$ consisting of $d+1$ elements. Fix $i \in \{1, \dots, q\}$ and let x_i satisfy

$$u_j^t x_i = 1 \quad \text{for all } j \in J_i. \tag{21}$$

Let R_i denote the set of $\varepsilon \in R^n$ for which the system of equations

$$u_j^t x = 1 + \delta\varepsilon_j \quad \text{for all } j \in J_i \tag{22}$$

is feasible. Using the fact that x_i satisfies (21) we see that R_i coincides with the set of $\varepsilon \in R^n$ for which

$$u_j^t x = \varepsilon_j \quad \text{for all } j \in J_i, \tag{23}$$

is feasible. Since x in (23) is constrained to lie in R^d, and there are $d+1$ variables ε_j, $j \in J_i$, the set of $\{\varepsilon_j, j \in J_i\}$ for which (23) has a solution forms a proper subspace of a $d+1$-dimensional space. It follows that R_i forms a proper subspace of R^n.

Let $E = \bigcup_{i=1}^q R_i$. If $\varepsilon \in R^n \setminus E$, and $\delta \in (0, \Delta)$, where Δ is given in (a), then none of the systems of equations (22) is feasible so $J_i \notin \mathscr{F}^{\varepsilon,\delta}$, for $i = 1, \dots, q$. Furthermore, by (a)(ii), $\mathscr{F}^{\varepsilon,\delta} \subseteq \mathscr{F}$ and it follows that $\mathscr{F}^{\varepsilon,\delta}$ does not contain any faces with $d+1$ elements, so $\mathscr{P}^{\varepsilon,\delta}$ is in general position. □

Acknowledgments. The Associate Editor and referees deserve much thanks for suggesting many improvements. We express our gratitude to Hal Sadowsky for simplifying our original proof of Theorem 3.

REFERENCES

Borsuk, K. (1948). On the imbedding of systems of compacta in simplicial complexes. *Fund. Math.* **35** 217–234.

Devroye, L. (1986). *Nonuniform Random Variate Generation*. Springer, New York.

Donnelly, T. G. (1973). Algorithm 462. Bivariate normal distribution. *Communications ACM* **16** 638.

Edelsbrunner, H. (1995). The union of balls and its dual shape. *Discrete Comput. Geom.* **13** 415–440.

Edelsbrunner, H. (1987). Algorithms in Computational Geometry. Springer, New York.

Edelsbrunner, H. and Seidel, R. (1986). Voronoi diagrams and arrangements. *Discrete Comput. Geom.* **1** 25–44.

FISHMAN, G. (1996). *Monte-Carlo*. Springer, New York.

FRIGESSI, A. and VERCELLIS, C. (1984). An analysis of Monte Carlo algorithms for counting problems. IAMI-84.2, Dept. Mathematics, Univ. Milan.

HILTON, P. (1960). *Homology Theory*. Cambridge Univ. Press.

HIRSCH, M. (1976). *Differential Topology*. Springer, New York.

MURTY, K. (1983). *Linear Programming*. Wiley, New York.

NAIMAN, D. and WYNN, H. P. (1992). Inclusion–exclusion-Bonferroni identities and inequalitiues for discrete tube-like problems via Euler characteristics. *Ann. Statist.* **20** 43–76.

NAIMAN, D. Q. and WYNN, H. P. (1993a). Independence number, Vapnik–Chervonenkis dimension, and the complexity of families of sets. *Discrete Math.* **154** 203–216.

NAIMAN, D. and WYNN, H. P. (1993b). Indicator identities via Voronoi diagrams. Unpublished manuscript.

NAIMAN, D. Q. and WYNN, H. P. (1993c). A theorem on independence. *Discrete Math.* **120** 287–289.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. and VETTERLING, W. T. (1988). *Numerical Recipes in C*. Cambridge Univ. Press.

ROTMAN, J. J. (1988). *An Introduction to Algebraic Topology*. Springer, New York.

TAKÁCS, L. (1967). On the method of inclusion and exclusion. *J. Amer. Statist. Assoc.* **62** 102–113.

TONG, Y. L. (1980). *Probability Inequalities in Multivariate Distributions*. Academic Press, New York.

DEPARTMENT OF MATHEMATICAL SCIENCES
JOHNS HOPKINS UNIVERSITY
BALTIMORE, MARYLAND 21218
E-MAIL: dan@jesse.mts.jhu.edu

DEPARTMENT OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCE
CITY UNIVERSITY
LONDON
ENGLAND